Scalability of the VSO

Virtual Solar Observatory

Underlying Complexity

Central Servers

VSO2
VSO1
Central server

MONITORING
STATUS

Data Providers

Data Repositories | Data Query Interfaces

MSU - SXT
SXT Archive | SXT Interface

NASA - SDAC
SOHO Archive | SOHO Interface

NOAO-NSO
NSO Archive | NSO Interface

Stanford - SHA
SHA Archive | SHA Interface
YOHKOH Flare Catalog | YOHKOH Flare Interface

UPDATES | CHECK FOR UPDATES | LOGGING

VSO Instances

Alt. Server
Public Server
Local Server
Registry
Core

QUERIES
RESPONSES
QUERY
RESPONSE

Access Methods

User Query Interfaces | User Interfaces

Access through VxO
VOTable UQI | VxO
Access through VSPO
VSPO UQI | VSPO
Access through EGSO
EGSO UQI | EGSO
Web based access
User Query Interface | Web based forms

Users

Solar Physicist (not to scale)

QUERY
RESPONSE

Perceived Simplicity

Data Providers

MSU Archive
SOHO Archive
NSO Archive
SHA Archive
YOHKOH Flare Catalog

VSO Instance | **Access Method** | **User**

QUERIES | QUERY | QUERY
RESPONSES | RESPONSE | RESPONSE

Abstract:

The successful development of the VSO prototype raises fundamental system architecture issues in dealing with the scalability and fault tolerance of distributed systems. The four sites participating in development (Stanford, MSU, NSO and SDAC) are only the starting point of VSO.

The system's overall architecture is being influenced by interaction issues between different data repositories that do not share a common set of capabilities. As new data providers or new search capabilities are added, the time needed to return useful results should not grow linearly with them. We take care to identify and isolate slow or unresponsive data repositories, to keep them from adversely affecting the overall quality of VSO searches.

This extra complexity needs to remain unobtrusive to the solar researches, so that they can focus on their research, and not the tools that they are using.

Monitoring:

To assure that users are given the most comprehensive search possible, in a reasonable amount of time, it is necessary to monitor the availability of individual components. If a component seems to be performing outside the range of accepted behaviour, one of the VSO Central Servers will determine its availability, and should it be unresponsive, will watch for it to return to an acceptable state.

Scalability:

The modular design used by VSO allows the addition of new features or functionality without impacting unrelated components. The design provides for horizontal scalability by allowing each VSO component, except for the individual Data Providers, to be implemented on more than one system, to distribute the workload.

Fault Tolerance:

By breaking the tasks down into smaller components, we create self contained items which are horizontally scalable. As the VSO design does not require the maintenance of session state on a component, we can use a failover design to provide high availability of VSO components without the overhead of a more complex clustered environment.

By maintaining information about components in the VSO Registry, we can easily add information about new Data Repositories, or temporarily skip over a Data Repository that might be down for maintenance or otherwise unavailable or unresponsive.
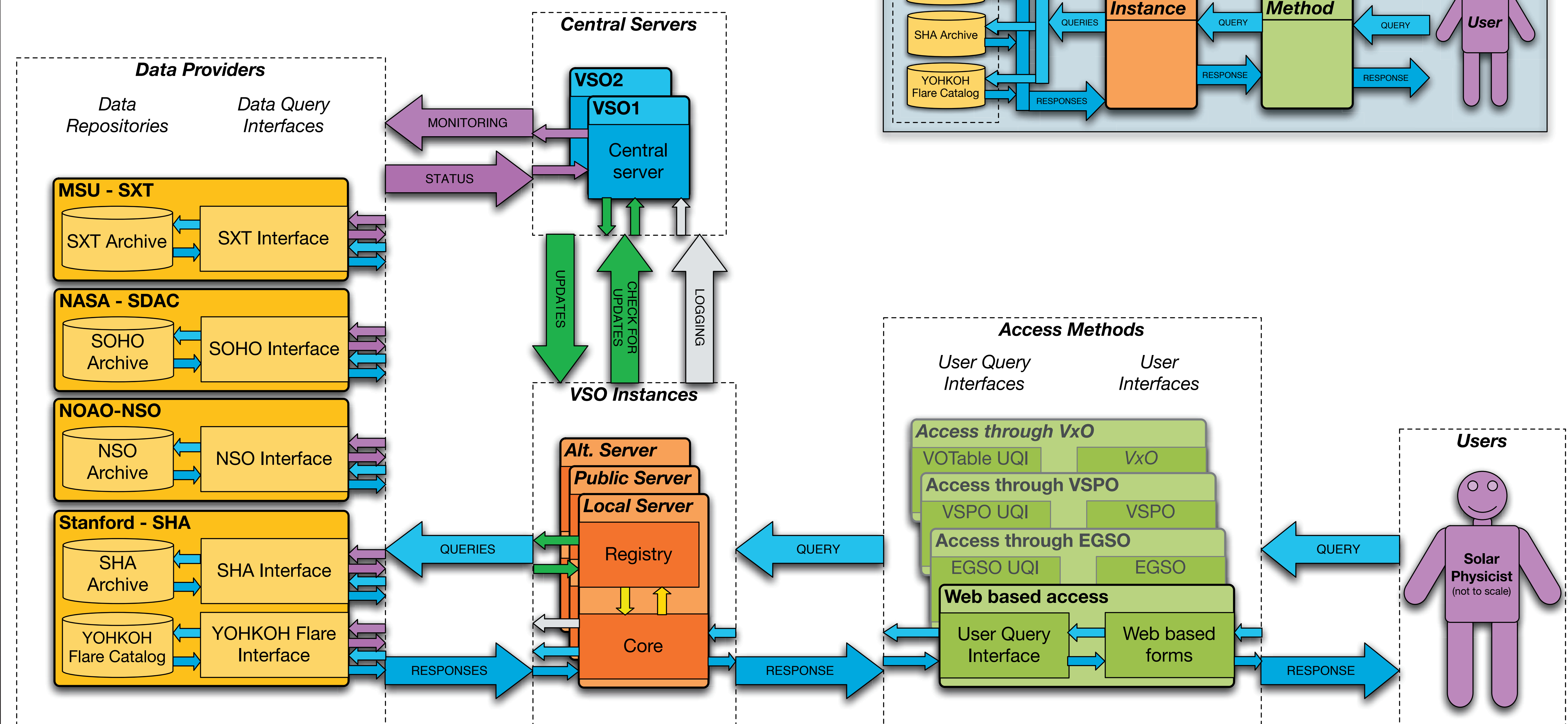
Modularity:

Each individual component of VSO performs simple tasks which provides the overall complex functionality an end user.

User Interface:
Handles the presentation to the user
User Query Interface:
Translates queries to the format used internally by VSO
VSO Core:
Handles the multiplexing of the queries to multiple data providers.
Registry:
Provides directory information about the different Data Providers.
Data Query Interface:
Translates VSO queries to a format that a Data Provider can understand.
**Data Repository:**
An online service containing solar data.
Central Server:
Handles distribution and collection of information between VSO instances.

Tuning:

The simple design of the VSO components allows for each component to be individually tuned to receive the best performance characteristics possible. The simplicity of the Registry design allows us to use vague descriptions to quickly get information available online, or to use more specific descriptions to reduce the number of false positives that might otherwise occur.
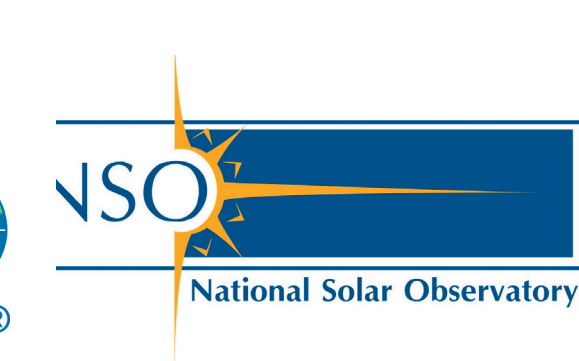
The centralized logging capabilities allow the operations staff to review the logs across multiple VSO instances to determine when and where false positives are occurring, and adjust the Registry entries for a Data Provider to reduce unnecessary queries.

NASA | SRI | NSO National Solar Observatory | Stanford | MONTANA STATE UNIVERSITY BOZEMAN
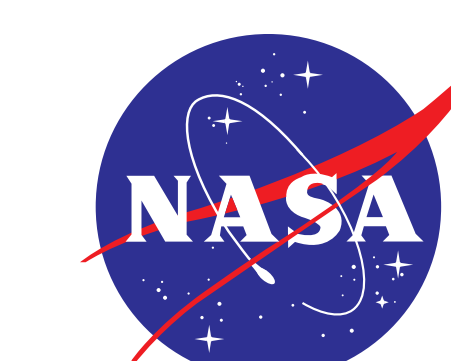
J.A. Hourclé (GSI/NASA GSFC) | A.R. Davey (SwRI) | I. Suarez-Sola (NSO) | K.Q. Tian (Stanford U.) | K. Yoshimua (MSU-Bozeman)